Promotion of Virtual Research Communities in CHAIN

Tiwonge BANDA[1], Roberto BARBERA[2,3], Bruce BECKER[4], Ludek MATYSKA[5], Rafael MAYO-GARCÍA[6,*], Margaret NGWIRA[1], Luis NÚÑEZ[7], Ognjen PRNJAT[8], Manuel RODRÍGUEZ-PASCUAL[6], Antonio Juan RUBIO-MONTERO[6], Federico RUGGIERI[9], Neeraj SINHA[10]

(On behalf of the CHAIN Project)

[1] *UbuntuNet Alliance, P.O. Box 2550, Lilongwe Malawi*
[2] *INFN Catania, Via S. Sofia 64, Catania, Italy*
[3] *University of Catania, Viale A. Doria 6, Catania, Italy*
[4] *Meraka Institute, CSIR Site, Meiring Naudé Road, Brummeria, Pretoria, South Africa*
10 *CESNET, Zikova 1905/4, 160 00 Praha 6 – Dejvice, Prague, Czech Republic*
10 *CIEMAT, Avda. Complutense, 22, 28040 Madrid, Spain,* [*]*Email: rafael.mayo@ciemat.es*
[7] *RedCLARA, Rambla República de México 6125, Montevideo 11400, Uruguay*
[8] *GRNET, 56, Mesogion Avenue, Athens, Greece*
[9] *INFN Roma Tre, Via della Vasca Navale 84, Roma, Italy*
[10] *Office of the Principal Scientific Adviser to the Government of India, 326, Vigyan Bhavan Annexe, Maulana Azad Road, New Delhi 110011, India*

Abstract

The CHAIN project has been working on coordinating and leveraging the efforts made over the past years to extend the European e-Infrastructure (and particularly the Grid) operational and organisational principles to a number of Regions over the world. In this sense, the project is currently working on elaborating a strategy and defining instruments in order to ensure coordination and interoperation of the European Grid Infrastructure with those emerging in other regions of the world (Asia, Mediterranean, Latin America and Sub-Saharan Africa). Among these efforts, the promotion of Virtual Research Communities plays a key role. In this work, the progresses made by CHAIN during these two years in the aforementioned topic are presented

Keywords

Grid; Virtual Research Communities; CHAIN.

1. Introduction

The effort provided by the EC over the last decade in e-Science has been focused not only in Europe, but also on different world regions, mainly Africa, Asia and Latin America. At the same time, different layers of e-Infrastructures, computing platforms and related human power have been of interest to the EC (Research and Education Networks, Grid and HPC infrastructures and Virtual Research Communities).

Due to the economic status in the aforementioned Regions some years ago, Grid emerged as the main actor for producing e-Science. Thus, the coordination of these world-wide efforts has been mostly pushed, but even in this case, such coordination has merely be restricted to basic

operational, organisational and technological know-how transfer/exchange, i.e. no big advances have been made until now to link e-Infrastructures at intercontinental level due to their specific requirements which depend on the targeted region.

The Co-ordination & Harmonisation of Advanced e-Infrastructures (CHAIN(2010), http://www.chain-project.eu) project, started on December 1st, 2010, has been working in coordinating and leveraging the efforts made over the past 6 years to extend the European e-Infrastructure (and particularly Grid) operational and organisational principles to a number of Regions in the world, mainly those identified for this purpose by the EC. Once that the project is coming to its end, it can be mentioned that this vision has been structured in several lines of performance: study of the state of the art of the current e-Infrastructures and e-Science status per region; needs and commonalities of these computing platforms; and, Virtual Research Communities (VRC) requirements and services.

Among many other actions carried out, CHAIN has pushed and supported the operation of the Africa & Arabia Regional Operation Centre (Africa ROC for short, http://roc.africa-grid.org/), which is crucial for a further development in the Region of academic, scientific and cultural advances based on ICT.

Besides, a wider knowledge Base about Grid infrastructures and scientific applications is available through the CHAIN webpage (http://www.chain-project.eu/knowledge-base)

CHAIN WP3, 'Present and emerging needs of trans-continental scientific communities', devotes its framework to the study of VRCs. As a first step, a coordinated data collection on the existing Grid state of the art across the world was carried out. Such a work was culminated with the collaboration with different communities that have supported the project in a symbiotic way. In addition, a study on the requirements and services of interest for the VRCs that had to be offered by the infrastructure providers has been documented by the project. This study has many specific actions, but all of them are based on three main pillars: sustainability, adoption of standards and a collaborative scenario.

Specifically, the actions that have been carried out in order to progress in the aforementioned topics have been:

- To deliver once a year a VRCs-related survey to the National representatives. Such a questionnaire was firstly implemented by CHAIN WP2 'Consolidation of existing state of the art'.
- To get a continuous feedback about the different versions of the road-map of services from the National representatives too.
- To maintain a continuous liaison with the identified VRCs fostering the collaboration by periodic contacts, which also include the reception of their feedback about the different versions of the road-map.
- To consolidate such collaboration with the VRCs by means of the signature of Memoranda of Understanding.
- To organise specific workshops that have been held as part of major conferences (http://agenda.ct.infn.it/categoryDisplay.py?categId=80) in order to maximise its impact and, on the other hand, to know the requirements and necessities that VRCs have and present them, at the same time, the different versions of the road-map.
- To attend several events where the work carried out to that moment has been also presented.

As a consequence of this activity, the results related to every associated VRC are presented below as well as the main basis adopted by CHAIN in order to improve the impact of Grid computing worldwide: the Science Gateway paradigm.

2. The Science Gateway paradigm

It has been demonstrated that using Grid is not straightforward. There can be found several barriers, but undoubtedly, one major drawback is the relatively high level ICT skills requirement for beginners and the way of accessing it (authentication & authorisation). Users have to cope with complex security procedures, execution scripts, job description languages, command line based interfaces and lack of standards. This makes the learning curve very steep and keeps non IT-experts away.

TeraGrid project defined a Science Gateway (SG) as a community-developed set of tools, applications and data that is integrated via a portal or a suite of applications, usually in a graphical user interface, that is further customized to meet the needs of a specific community. This way, it will be possible to abstract the final user from the technological complexity that would be underneath.

The second step is to get an easier access to the SG. The adoption of robot-certificates for managing the jobs to be executed and the authorisation to get into the SG by means of Identity Federations (see for example eduGAIN), which count on millions of users (so, for example, any researcher belonging to a R&D Centre or to a University can immediately login by his own credentials), is an outstanding asset for this paradigm.

Last action is then to develop a computational framework where available and new developments could be coupled (added) as independent linked modules. Doing so, it will be easier to integrate new applications and codes in the framework. Such an implementation has been carried out at INFN-Sezione di Catania.

Because of all these reasons, the CHAIN Science Gateway (http://agenda.ct.infn.it/conferenceDisplay.py?confId=794) is supposed to provide a short term answer to interoperability when several different e-Infrastructures have to cooperate in order to fulfil the requirements of intercontinental VRCs and is being proposed and fostered by CHAIN.

The framework for SG is fully web-based and adopts official worldwide standards and protocols, through their most common implementations. These are:

- The JSR 168 and JSR 286 standards (also known as "portlet 1.0" and "portlet 2.0" standards);

- The OASIS Security Assertion Markup Language (SAML) standard and its Shibboleth and SimpleSAMLphp implementations;

- The Lightweight Direct Access Protocol and its OpenLDAP implementation;

- The Cryptographic Token Interface Standard (PKCS#11) standard and its Cryptoki implementation; and,

- The Open Grid Forum (OGF) Simple API for Grid Applications (SAGA) standard and its JSAGA implementation.

In addition, SG is built using the Liferay portal framework and can be downloaded and installed by a virtual machine containing the development environment and examples of basic template portlets that can be customised to integrate specific applications. The CHAIN web page offers continuous updated information about all these topics on the "Applications" section (http://www.chain-project.eu/applications).

The SC paradigm has been the main actor in the recently CHAIN worldwide interoperability demo (http://science-gateway.chain-project.eu/) that was showed during the EGI Technical Forum 2012 in Prague in September 2012. In the demo, several contributors were present in addition to CHAIN:

- e-Science projects/initiatives: CNGrid; EGI-InSPIRE; EUMEDGrid Suuport; Future Grid; GARUDA; GISELA; SAGrid; and, WeNMR.

- Middleware developers: EMI; GENESI II; gLite; GOS; JSAGA; OurGrid; and, UNICORE,

- ROCs: Africa & Arabia ROC; EGI associated ROCs; IGALC; ROC-LA

- Applications that could be executed: ASTRA; GROMACS; jModelTest; Octave; Parallel 'Hello world'; ProtTest3; R; Sequential 'Hello world'; and, Sonification

The demo was a success and it was able to run different kind of jobs on different infrastructures by means of different middleware. In addition, and as a key point, with a user's complete unattended control, who simply selected the input and submitted the job by clicking on a "Run" icon.

It is worth mentioning that the CHAIN worldwide interoperability demo addressed most of the recommendations proposed by the VRCs (see Section 4).

3. The CHAIN collaborative Virtual Research Communities

This section describes the scientific and technical communities that have been selected by CHAIN in order to profit from join collaboration and, as final goal, to propose a possible model of services to be implemented by the e-Infrastructures.

They have different levels of maturity in what Grid development concerns and are formed on different basis ranging from well-established communities even funded by external entities to a group of users who employ a common application or have a synergy in their scientific interest. This heterogeneity is as an asset for CHAIN in order to propose a new portfolio of services that could really fit a broad range of groups. Even more, their ICT interests move from applied Science to Humanities, so researchers with different scopes can be addressed. A risk actually lies in this approximation, but only acting on similar and well developed Grid initiatives (either from the same scientific area or not) would make the final proposed model useless for many people and/or e-Infrastructures providers. In addition, they have also been selected because of their sustainability and spread continental presence, which is a must for CHAIN.

During the two years lifetime of CHAN, there has been a continuous exchange of information between the associated VRCs and the project. In general terms, it could be mentioned that CHAIN has provided the VRCs with:

- New users all around the world;

- New codes of interest to the communities;

- Access to infrastructures by means of the interoperability demo;

- Access to training activities with the collaboration and support of EPIKH project (see the CHAIN training agenda at http://agenda.ct.infn.it/categoryDisplay.py?categId=81); and,

- Information about the different actions that have been running worldwide, mainly in the Regions targeted by CHAIN.

On the other hand, the VRCs have provided CHAIN with very valuable information in:

- Regional presence and impact of the different communities;

- Best practices and requirements; and,

- Feedback about the CHAIN published information.

In a equidistant point or, in other words, a two-way channel, it can be mentioned the relationship with EGI and EGI-InSPIRE, since we have tried to coordinate all the challenges we had in accordance to the European initiative and, at the same time, VRCs have made their best in order to produce their e-Science in this defined technological framework.

In the following subsections, the order in which the VRC are presented is chronological on a first-contacted first-served basis.

3.1 We-NMR

We-NMR (http://www.wenmr.eu/) is a project which aims to optimize and extend the use of the NMR and SAXS research infrastructures through the implementation of an e-Infrastructure in order to provide the user community with a platform integrating and streamlining the computational approaches necessary for NMR (Nuclear Magnetic Resonance) and SAXS (Small Angle X-ray Scattering) data analysis and structural modelling. When CHAIN was started, We-NMR had presence in Nigeria and South Africa in the continent.

A MoU was signed with this initiative on September 2011 and its presence has been continuous in the workshops organised by CHAIN during the reporting period as a continuation of the first contacts held in the KoM. Several discussions have been held with the coordinators of WeNMR and their feedback has been very useful for the drawing-up of the WP3 CHAIN deliverables. Once these documents were published, they have been presented in several forums, but the summary of its recommendations have been directly consulted with WeNMR in order to improve the quality of such recommendations to appear in future deliverables.

By means of the CHAIN WP2 survey, new contacts of researchers interested in the areas covered by WeNMR were provided to the VRC. The countries where these scientists were settled were Burundi and Costa Rica. WeNMR counts on an own-developed web access to submit jobs to the Grid, but discussions have been held to test and adopt the SG paradigm or benefit from its methodologies in order to improve the WeNMR submission.

3.2 WRF4G

WRF4G (http://www.meteo.unican.es/es/software/wrf4g) is a Grid version of the well-known Weather Research and Forecasting (WRF) modelling system application. It is widely used by the meteorological agencies and many other groups in the Earth Science domain. Its Grid-based version has increased the resources where tasks such as Idealized simulations, Regional and Global applications, Parameterization and Data assimilation research or Forecast and hurricane

research can be performed. When CHAIN was started, WRF4G had presence in Ghana, Senegal and South Africa in the continent.

A MoU was also signed with this initiative on September 2011 and its presence has been continuous in the workshops organised by CHAIN as a continuation of the first contacts held in the Kick-off Meeting. Several meetings have been held with the coordinators of WRF4G and their feedback has been useful in drawing the WP3 deliverables. The most important information of them has been also directly consulted with WRF4G representatives in order to improve the quality of recommendations to appear in future releases.

By means of the CHAIN WP2 survey, new contacts of researchers interested in the scientific applications of WRF4G were provided to the VRC. The countries where these scientists were settled were Burundi, China, Costa Rica and Cuba.

WRF4G has been working on a similar basis in order to port to the Grid the CAM model, so a new Grid release called CAM4G is available. CHAIN has also supported this new version and has disseminated it in the events it has organised. A direct liaison between the WRF4G initiative and the Climate Change community is being supported by CHAIN (see Subsection 3.6).

WRF4G relies on a command line method for the submission of jobs, so its incorporation to the SG paradigm has been foreseen. Nevertheless, since WRF4G is mostly based on GridWay metascheduler (which distributes tasks by means of DRMAA) and SG works with SAGA, an actor for communicating both methodologies should be developed. WRF4G has kept disseminating CHAIN plans and outcomes inside the Climate Change CORDEX project.

3.3 jModelTest / ProtTest3

These two applications belong to the Life Sciences domain and, in particular, to the Evolutionary Biology. Both are freely available on-line (http://darwin.uvigo.es/) for the statistical selection of best-fit models of nucleotide substitution (jModelTest) and amino-acid (ProtTest3) replacement for a given set of aligned sequences. Thus, many researchers interested in molecular systematics, phylogenetics, phylogenomics, molecular evolution and/or bioinformatics use them continuously all around the world.

ModelTest has around 30,000 registered users worldwide while jModelTest (the Java version) has around 11,000 and ProtTest around 5,000, i.e. almost all countries in the world have scientists who are using one of these tools, African ones too.

CHAIN signed a MoU with this initiative on February 2012 and has also supported it by implementing distributed versions of both codes that can be executed either on local clusters or Grid. Thus, these versions have been integrated in the SG paradigm in both their sequential and distributed releases (http://gisela-gw.ct.infn.it/jmodeltest and http://gisela-gw.ct.infn.it/prottest) by means of an actor between DRMAA and SAGA.

By means of the CHAIN WP2 survey, new contacts of researchers interested in the scientific applications of jModelTest and ProtTest3 were provided to the VRC. The countries where these scientists were settled were Burundi, Costa Rica, Democratic Republic of Congo, Ethiopia, Nigeria, Panama, Sudan and Taiwan.

3.4 INDICATE

This project (http://www.indicate-project.eu) is working on coordinating policy and best practices regarding the use of e-Infrastructures for Digital Cultural Heritage. The project aims at

establishing and stimulating a network of common interest made up of experts and researchers in all the relevant fields, whose sustainability will be planned on a long term beyond the project lifetimes. Egypt was already collaborating with the project when CHAIN was started.

INDICATE focuses its activities in countries all around the Mediterranean, both European and African. Nevertheless, contacts have been established with China and Latin American. CHAIN has signed a MoU with INIDCATE on March 2012.

Humanities have a very different approach to ICT than Science. In addition, their interests and required services are diverse, which improves the quality of the road-map and validation tests to be done by CHAIN. In this way, new contacts to integrate a digital repository in Panama by means of the INDICATE e-Culture SG have been provided. Since digital repositories are becoming more and more important, the collaboration between INDICATE and CHAIN is expected to gain from this fostered activity.

As in the previous cases, the different documents produced by CHAIN have been sent to the INDICATE coordinator in order to get her feedback.

3.5 DECIDE

DECIDE (http://www.eu-decide.eu/) objective is to design, implement, and validate a Grid-based e-Infrastructure building upon neuGRID and relying on the Pan-European backbone GEANT and the NREN. Over this e-Infrastructure, a service will be provided for the computer-aided extraction of diagnostic markers for Alzheimer's disease and schizophrenia from medical images.

A MoU with this VRC was signed on April 2012, but a close collaboration has been established anyway during the whole lifetime of CHAIN by common partners shared between the two initiatives, the direct contacts between the managerial bodies (road-map of services has been sent to the VRC coordinator requesting his feedback) and the participation of DECIDE in the events organised by CHAIN. In addition, the project has provided DECIDE with new contacts in Nigeria.

As a result, the SG paradigm promoted by CHAIN has been also adopted by DECIDE and the community has participated with their own web-access in the validation model that CHAIN is proposing.

3.6 Climate Change

A Climate Change-related VRC has been promoted by CHAIN since the very beginning (see the conference on the "Role of e-Infrastructures for Climate Change research" at http://users.ictp.it/~smr2238/). This topic is of outmost importance nowadays and has researchers interested in all around the world, so the work has been focused on a two-fold basis: looking for well-established European projects that were aware of the ICT advances in what Climate Change refers and looking for researches in the Regions of interest to CHAIN who could be interested in joining either the developments offered by the European initiatives or form a new community around a well-known application. In this sense, the dissemination and promotion of WRF4G among these scientists have been carried out by CHAIN.

As a result, CHAIN has got the interest from researchers settled in ten (10) non-European countries (Algeria, Cameroun, Cote d'Ivoire, Ethiopia, Morocco and South Africa from Africa) and has identified four (4) initiatives that could help as a starting point to definitively form a VRC. In most of these groups and with the support of ICTP, the most used application is RegCM. In addition, CHAIN has also offered WRF4G code developers support to work on the

applications used by these researches in order to port them to the Grid with the same methodology applied to CAM4G, the action of which could drive to the creation of a VRC too. Logically, both strategies could find synergies to meet and, due to this, the new survey sent to the National representatives raise this point.

Last, groups devoted to the study of seismology have been identified in Latin America with the support of the GISELA project (they are interested in porting to the Grid the SPECFEM3D code) and India through of ICTP.

3.7 LSGC

Due to its wide presence, the Life-Science Grid Community (LSGC, http://lsgc.org/en/LSGC:home;jsessionid=D1E7996F7665C3FA900B13A2116E1A36) has been approached by CHAIN. LSGC is one of the biggest Grid users communities, HEP excluded, and counts on a solid background in this computational platform, where codes of different characteristics are being utilised nowadays. Besides, according to EGI accounting portal, during 2012, there have been jobs running in these regions: Asia Pacific, Canada, Europe, Latin America and Russia.

LSGC representatives have been invited to make presentations about their communities from the first event organised by CHAIN and a MoU was signed between this initiative and CHAIN in March 2012.

Since LSGC is mature enough and has very well established links with the people who participate in it, CHAIN has worked in finding new users in the targeted Regions and, at the same time, has got the initiative feedback for the published documents profiting from its broad expertise.

3.8 SuperB

The SuperB flavour factory (http://superb.infn.it/home) is a major international research centre for fundamental and applied physics to be built on the campus of the University of Rome Tor Vergata. It is an accelerator that will provide complementary information to LHC, looking at rare decays with a very high luminosity electron-positron asymmetric collider. In Africa, there have been contacts in Algeria, Egypt, Ghana, Madagascar, Morocco, Mozambique, Rwanda, South Africa and Tunisia from the very beginning.

From the CHAIN point of view, it is a novel approach to the research communities that could deserve very good results because it is based on the major Grid users group, i.e. HEP, but at the same time, is not limited to the rules and policies that the four ICT Divisions belonging to the LHC experiments impose or recommend due to its huge Grid use. In this way, groups interested in this kind of physics, but devoted to other kind of problems, can offer a new scenario and, besides, can collaborate with SuperB from their own place. SuperB VO is currently supported by five (5) National Grid Initiatives, none of them belonging to the Regions of interest to CHAIN.

Coming from outside Europe, groups belonging to twenty three (23) countries have been identified and the first discussions with the SuperB representatives have been carried out in order to sign a new MoU which better defines a strategy for collaboration. In the meantime, CHAIN is also looking for new contacts.

4. Some hints about the proposed road-map of services

CHAIN has been working on drawing up a road-map of services and requirements that are needed by the VRCs to be proposed to the Distributed Computing Infrastructures (DCI).

The information is divided in two main areas: the current status of DCIs in what VRCs concerns; and, the road-map itself. In addition, the latter is divided into 88 itemized recommendations grouped in Technical (64), Training, administration and use (14) and Collaborative scenario (10) ones.

The whole and latest list of such recommendations can be found in WP3 deliverable D3.4 'Road-map of trans-continental e-infrastructures for virtual communities Updated' at http://www.chain-project.eu/status.

5. Conclusions

The CHAIN project has been working on coordinating and leveraging the efforts made over the past years to extend the European e-Infrastructure (and particularly the Grid) operational and organisational principles to a number of Regions over the world. In this sense, the project keeps working on elaborating a strategy and defining instruments in order to ensure coordination and interoperation of the European Grid Infrastructure with those emerging in other regions of the world (Asia, Mediterranean, Latin America and Sub-Saharan Africa). Among these efforts, the promotion of Virtual Research Communities plays a key role.

CHAIN has strongly collaborated with eight different VRCs and has promoted their interests in the previously mentioned regions. On the other side, CHAIN has profited from the VRCs' experience for documenting a road-map of services to be provided by the DCIs. Such a road-map has been successfully tested by means of the CHAIN worldwide interoperability demo, which had the Science Gateway paradigm as major actor. In the demo, it was demonstrated that it is possible to run different kind of jobs on different infrastructures by means of different middleware in a complete unattended way.

Reference

CHAIN, (2010) Co-ordination & Harmonisation of Advanced e-Infrastructures EU FP7 project http://www.chain-project.eu) project